Analysis of Roles and Groups in Blogosphere

Bogdan Gliwa, Anna Zygmunt, and Jarosław Koźlak

Abstract. In the paper different roles of users in social media, taking into consideration their strength of influence and different degrees of cooperativeness, are introduced. Such identified roles are used for the analysis of characteristics of groups of strongly connected entities. The different classes of groups, considering the distribution of roles of users belonging to them, are presented and discussed.

1 Introduction

In recent times an increasing popularity of social media, for example blogosphere, has taken place. It has a more and more increasing importance for marketing forecasting or predicting the popularity of persons and ideas. Dynamic interactions between persons may influence the development of groups of strongly connected entities. Nowadays, applying of the methods of social network analysis is a popular approach for understanding the behaviour of such systems.

The aim of this work was the identification of influential entities and ranges of their influence both on the global and local (in the scope of groups) level. In the analysis, it is assumed that the influence may be built by taking part in discussions regarding personally proposed subjects or by active participation in discussions of subjects proposed by others.

Such analysis may be conducted by recognition of characteristics of existing groups, identification of different kinds of groups by taking into account the presence of the users building their position or taking part in the discussion of other subjects proposed by others. This helps in drawing conclusions concerning the methods of gaining popularity, kinds of influences, their stability and duration.

Bogdan Gliwa · Anna Zygmunt · Jarosław Koźlak
AGH University of Science and Technology
Al. Mickiewicza 30, 30-059 Kraków, Poland
e-mail: {bgliwa,azygmunt,kozlak}@agh.edu.pl

2　Overview

Roles in blogosphere. There are many definitions of concept of role [12], [3], [13]. In social media, *role* can be treated as a set of characteristics that describe behavior of individuals and their interactions between them within a social context [7]. A lot of studies relate to certain social media and attempt to define their specific roles [9]. For example, an analysis of the basic SNA measures has been used in several studies to define social roles of *starters* and *followers* in blogosphere [6], [8]. *Starters* receive messages mostly from people who are well-connected to each other, and therefore they can be identified by low in-degree, high out-degree and high clustering coefficient in the graph. The distinction between the roles is obtained by combining the difference between the number of in-links and out-links of their blogs.

Groups. Usually it is assumed that the group is a set of vertices which communicate to each other more frequently than with vertices outside the group. Many methods of finding groups have been proposed [2], taking into account the time and its impact on the life cycle of the groups [1] [11]. In [14], [5] algorithm SGCI (Stable Group Changes Identification) was proposed, which we use mainly with CPM (Clique Percolation Method) [10]. The algorithm consists of four main steps: identification of short-lived groups in each separated time interval; identification of group continuation (using modified Jaccard measure), separation of the stable groups (lasting for a certain time interval) and the identification of types of group changes (transition between the states of the stable groups).

3　General Model

The analysed social system is characterised by the following elements: global roles played by given entities in the whole network, groups identified in a network of strongly interacting entities and roles assigned to entities in each group. The analysis of the considered social system is presented and the mentioned elements are calculated in time steps.

Global roles. Global roles present activity and influence of given entities in the scope of the whole social system. They distinguish also the specific features which characterise the behaviour of the entities: post influence, comments influence, local commenting in the context of its own posts or commenting of posts belonging to other users or comments to them. Posts or comments may be considered as influential, which cause a significant response from other bloggers, expressed by numerous comments related to them.

Such a distinction comes from the fact, that there are two kinds of influential entities: (i) the ones who build only their own position and refer only to these users who comment on their posts or comments and (ii) more social ones who take part in other discussions and comment also the posts of other bloggers. Among such influential users it is possible to distinguish also such ones who write both influential

posts and comments (called later *User*) and such who write only the influential posts (called later *Blogger*).

Considering the above criteria, it is possible to distinguish the following important roles: (*Selfish Influential User*) – they build their own position, write influential posts and comments, mainly in the context of their own posts. (*Social Influential User*) – participate also in the contexts of posts written by other users, write influential posts and comments, (*Selfish Influential Blogger*) – they write only influential posts, their comments relate mostly to the context of their own posts, (*Social Influential Blogger*) – they write only influential posts, they are concerned in commenting both their own and other posts.

Among the remaining users it is possible to distinguish those who are no authors of significant posts, but actively comment on posts of others. They are: *Influential Commentator*, writing influential comments and *Standard Commentator*, writing sufficiently numerous, but not necessarily influential comments, and seldom – posts. Users writing very few posts and comments are classified as *Not Active* users.

The rest of the users, who are neither distinguished by the particular activity nor stopped entirely their activity on the portal, have been assigned to the role of *Standard Blogger*.

In the presented social system, especially important is the presence of the influential roles, to which the first four are assigned to.

Stable groups. In each time step the groups are identified using the algorithm for identifying overlapping groups. Then, their structure is compared with the structure of the groups existing in the previous and next step. If a given group may be considered as a continuation of the group from the previous time steps and such continuation exists at least during a given time interval, such a group can be treated as a stable group. The following elements are presented for stable groups.

Roles in groups. For each stable group and each time step, roles of entities in these groups are identified in the similar manner as the identification of roles for the whole social system. The given users in the given time steps may be described by the role played by them globally and by roles played in different groups they belong to. It is also possible to analyse the given users globally from the point of view of the whole considered time period (containing a given number of time steps) and determine the number of roles the user played in this time period globally or in the given groups.

Group characteristics. The groups may be characterised on the basis of percentages of the users having special characteristics, expressed by the sets of roles. They may be as follows:

- users with a high significance (*Selfish Influential User*, *Social Influential User*, *Selfish Influential Blogger*, *Social Influential Blogger*,*Influential Commentator*),
- cooperative users, (*Social Influential User*, *Social Influential Blogger*, *Influential Commentator*),
- users oriented mainly to building their own position in the network (*Selfish Influential User*, *Selfish Influential Blogger*).

Definitions. Let us define *Post Response* (pr_{post}) that it is a number of comments for a given post excluding the author's comments in his own thread. We are using the following notation $c(X, cond)$ that means the number of elements in set X that each element of the set fulfills condition *cond* and $c(X)$ means the number of all elements in set X. Furthermore, notation $posts_a$ denotes posts of author a, $comments_a$ - comments of user a and $w(cond)$ - expression that returns 1 when the condition *cond* is satisfied, otherwise this expression returns 0.

Post Influence ($PostInf_a$) describes how influential posts the author a writes, and can be defined as follows:

$$PostInf_a = \sum_{(i,j)} i \cdot c(posts_a, pr \geq j) - \sum_{(k,l)} k \cdot c(posts_a, pr < l) \qquad (1)$$

where i,k – weights; j,l – thresholds of influence necessary for assigning a given weight.

Let us define *Comment Response* (cr_a) for author a as a number of comments that refers to comments of author a. Using this term we can formulate *Comment Ratio* ($comRatio_a$) for author a (if the author a wrote at least one comment, otherwise *Comment Ratio* equals 0):

$$comRatio_a = cr_a / comments_a \qquad (2)$$

Comment Influence ($ComInf_a$) describes the impact of comments written by author a:

$$ComInf_a = \sum_{(i,j)} i \cdot w(comRatio_a \geq j) - \sum_{(k,l)} k \cdot w(cr_a < l) \qquad (3)$$

where i,k – weights; j, l – thresholds.

Comments that author a writes in his own posts are marked as $ownCom_a$. Using this notation we can also define *Comment Ego* ($ComEgo_a$) as:

$$ComEgo_a = ownCom_a / comments_a \qquad (4)$$

4   Description of Experiments

Data set. The analysed data set contains data from the portal *salon24*[1]. The data set consists of 26 722 users (11 084 of them have their own blog), 285 532 posts and 4 173 457 comments within the period 1.01.2008 - 31.03.2012. The presented results were conducted on half of this dataset - from 4.04.2010 to 31.03.2012. The analysed period was divided into time slots, each lasting 7 days and neighboring slots overlap each other by 4 days. In the examined period there are 182 time slots. In each slot we used the comments model, introduced by us in [4] - the users are nodes and relations

---

[1] Mainly focused towards politics, www.salon24.pl

between them are built in the following way: from user who wrote the comment to the user who was commented on or if the user whose comment was commented on is not explicitly referenced in the comment (by using @ and name of author of comment) the target of the relation is the author of post.

Role definition. The measures described in the model (section 3) in experiments have the following values:

Post Influence of author a is calculated as follows:

$$PostInf_a = 4c(posts_a, pr \geq A_p) + 2c(posts_a, pr \geq B_p)$$
$$+ c(posts_a, pr \geq C_p) - c(posts_a, pr < D_p) - 2c(posts_a, pr < E_p) \qquad (5)$$

where A_p, B_p, C_p, D_p and E_p are parameters, describing the strength of the influence assigned to user, when it exceeds the given threshold.

For global roles the following values were used: A_p=100, B_p=100, C_p=50, D_p=6 and $E_p = 1$. For local roles the parameters depend on the size of group: A_p=10*$\sqrt{size}$, B_p=$A_p/2$, C_p=$B_p/2$, D_p=0 and E_p=1.

Comment Influence for author a is calculated as follows:

$$ComInf_a = 4w(comRatio_a \geq A_c) + 2w(comRatio_a \geq B)$$
$$+ w(comRatio_a \geq C_c) - w(cr_a < D_c) - 2w(cr_a < E_c) - 4w(cr_a < F_c) \qquad (6)$$

where A_c, B_c, C_c, D_c, E_c and F_c are parameters.

In experiments we used the following values (constant for global and local roles): A_c=1.25, B_c=1, C_c=0.75. For global roles we set parameters D_c=50, $E_c = 20$ and F_c=10, but for local roles the parameters have values as follows (depending from group size): D_c=2*$\sqrt{size}$, E_c=$D_c/2$, F_c=$E_c/2$.

Using terms defined above, we can assign users into one of the following categories:

1. *Influential User* (*infUser*): $PostInf > 2$ and $ComInf > 0$

   a. *Selfish Influential User*: $ComEgo \geq 0.75$
   b. *Social Influential User*: $ComEgo < 0.75$

2. *Influential Blogger* (*infBlog*): $PostInf > 2$ and $ComInf \leq 0$

   a. *Selfish Influential Blogger*: $ComEgo \geq 0.75$
   b. *Social Influential Blogger*: $ComEgo < 0.75$

3. *Influential Commentator* (*infComm*): $ComInf > 0$ and $PostInf \leq 2$
4. *Standard Commentator* (*comm*): $c(comments) \geq 20$ and $c(posts) \leq 2$
5. *Not Active* (*notActive*): $c(posts) < 1$ and $c(comments) < 2$
6. *Standard Blogger* (*stdBlog*): User that does not match to any from above roles.

Groups. To extract groups from networks we used CPMd from CFinder[2] tool, for k equals 5. For group evolution, we took advantage of the SGCI method. Figure 1a

---

[2] www.cfinder.org

presents the overall numbers of groups summed up in all time slots and figure 1b shows how the number of groups varies in each time slot. One can see that groups at size 5 outnumber others (overall and in each time slot).
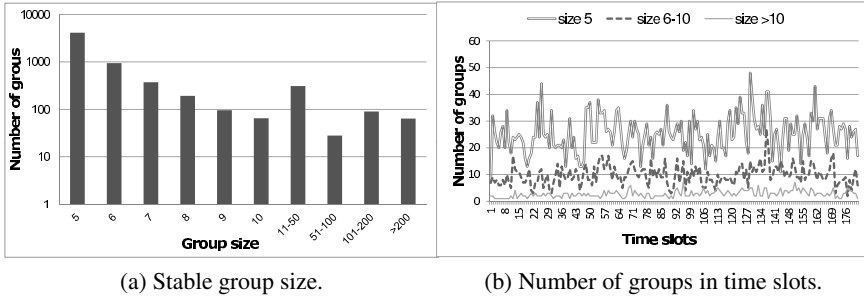

| (a) Stable group size. | (b) Number of groups in time slots. |

Fig. 1 Summary of stable groups

Roles. In figures 2a and 2b we can see how the roles are evolving during time respectively in whole network (global roles) and in groups (local roles). Both in network and in groups, the *Standard Bloggers* and *Commentators* represent the largest part among all roles. Furthermore, in the whole network there are many inactive users, which is very rare in local roles (a user can be inactive in the given time slot and be a member of a group when in a given time slot where he wrote nothing but he received some comments on what he had written in previous time slots). In whole network among important roles the most numerous are *Influential Commentators*. As we can see in figure 2a the percentage of *Influential Commentators* increases during time - it may have connection with the popularising using '@' in title of comment among users to indicate the response (we showed this trend in paper [4]). In figure 2b we can notice that in groups there are more *Influential Bloggers* than *Influential Commentators* (in contrast to figure 2b) which can suggest that most of the *Influential Commentators* are outside of these groups.

Tab. 1 presents global and local roles for some users placed high in ranking for each global role. We can observe that users who have global role *Influential User Selfish* mainly have local role *Influential Blogger Selfish*. Moreover, *Influential User Selfish* on a global level becomes *Influential Blogger Social* locally. Users that are *Influential Bloggers* globally (both *Social* and *Selfish*) are in the local level also as *Influential Bloggers*. But users playing global roles of *Influential Commentators* in very rare cases have the same role on the local level. It may indicate that influential commentators receive the most responses to their comments from outside their own groups or that these users write a lot of comments outside their group.

Roles in groups. We considered how large part of the groups is constituted by users with specific roles (role defined on global and local levels). Results can be observed on figures 3a and 3b. The biggest part of the groups constitute not important users (*Standard Bloggers* and *Commentators*), the largest groups (above 100 members) have smaller percentages of important users (both global and local
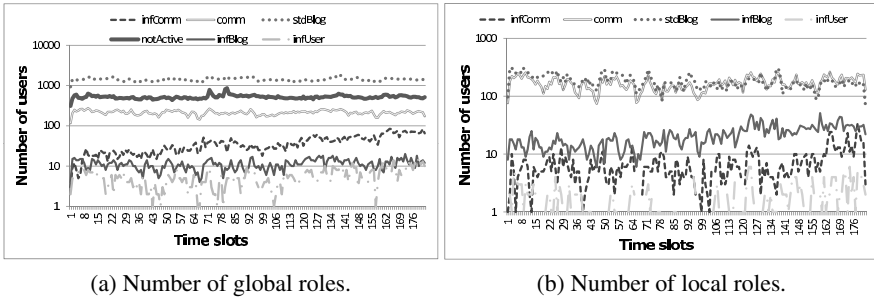
(a) Number of global roles.

(b) Number of local roles.

Fig. 2 Global and local roles in time slots

Table 1 Roles for selected users (how many times the user had given role, number in brackets means position in ranking)

UserId	Roles	infUserSel	infUserSoc	infBlogSel	infBlogSoc	infComm
2177	global	64(1)	5	14	1	19
	local	23	1	61	4	4
1672	global	41(3)	54(1)	27(15)	3	21
	local	11	19	58	39	14
241	global	41(3)	2	13	0	35
	local	12	1	44	2	12
796	global	13(10)	30(3)	9	17(4)	27
	local	5	6	19	42	4
11	global	1	46(2)	0	38(1)	47(13)
	local	0	1	1	65	0
657	global	0	0	141(1)	1	0
	local	0	0	239	2	0
783	global	3	1	94(2)	1	2
	local	1	0	134	7	0
1991	global	0	0	0	24(2)	0
	local	0	0	0	23	0
7325	global	0	4	0	0	79(1)
	local	0	0	0	9	8
549	global	4	4	7	1	69(2)
	local	0	0	17	41	2

levels) than smaller groups. It may mean that it is easier to play important roles in smaller groups.

Classes of groups. We considered small and large groups separately and look into how many users with given roles are inside them. We decided to take into consideration groups of size 9 as representative of small groups and groups at size between 150 and 250 as representative of large ones. Both cases have similar numbers of groups (there are 96 small groups and 85 large groups). We assigned groups to some classes based on 3 dimensions defined by the percentage of groups that 3 characteristics of role take (*Influential*, *Cooperative* and *Selfish* users). Using distribution of measures presented on figures 4a and 4b we defined threshold levels (for each dimension there are 2 thresholds) showed in tables 2a and 2b. Thresholds were defined by dividing measure distribution into 3 intervals with similar range. For each
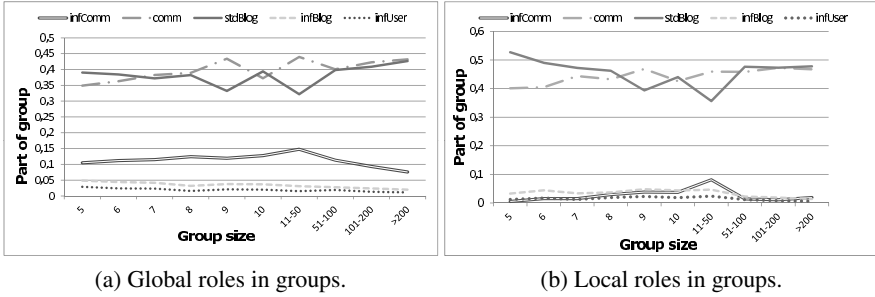
| (a) Global roles in groups. | (b) Local roles in groups. |

Fig. 3 Global and local roles in groups

dimension the ranges are ordered from 1 to 3 (label 1 concerns range below 1st threshold, label 2 - between 2 threshold, and 3 - above 2nd threshold).

In the results we obtained some classes for each case - tables 3a and 3b. We show the density for classes of small groups and the cohesion for large groups. One can see connections between these measures and classes of groups, for each class we calculated the mean value of density or cohesion for groups in that class. We can notice that more cooperative groups have higher values of density or cohesion for small and large groups (both on a global and local level). In the small groups on the global level, groups with the largest part of selfish roles are less dense.
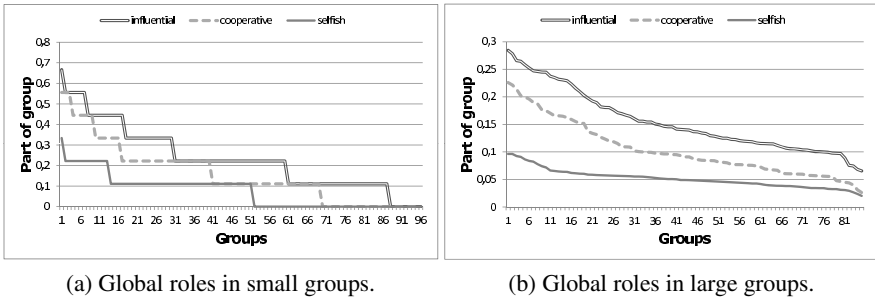

| (a) Global roles in small groups. | (b) Global roles in large groups. |

Fig. 4 Global roles in small and large groups

Table 2 Thresholds for classes

| (a) Small groups. | (b) Large groups. |

Roles	Influential	Cooperative	Selfish		Roles	Influential	Cooperative	Selfish
global	0.2,0.4	0.1,0.3	0.1,0.2		global	0.1,0.2	0.1,0.15	0.05,0.07
local	0.15,0.3	0.1,0.2	0.1,0.15		local	0.04,0.08	0.02,0.05	0.02,0.04

Table 3 Classes for groups (in rows: global/local)

(a) Classes for small groups.					(b) Classes for large groups.				
Influen.	Cooper.	Self.	Density	Count	Influen.	Coop.	Self.	Cohesion	Count
1	1	1	0.596/0.572	9/34	1	1	1	5.98/6.05	9/20
1	1	2	0.567/0.569	2/19	1	1	2	5.68/5.92	1/9
1	2	1	0.6/0.588	14/6	1	2	1	-/6.1	-/3
2	1	3	0.555/0.514	5/1	2	1	1	6.21/-	25/-
2	2	1	0.618/-	14/-	2	1	2	6.28/6.07	12/9
2	2	2	0.61/0.604	17/14	2	1	3	5.97/6.8	6/1
2	2	3	0.551/-	3/-	2	2	1	6.57/6.69	8/7
2	3	1	0.559/-	4/-	2	2	2	6.42/6.61	4/15
3	2	3	0.511/0.736	5/1	2	2	3	-/6.37	-/4
3	3	1	0.635/0.608	4/4	2	3	1	6.44/7.19	1/2
3	3	2	0.646/0.644	8/11	3	2	3	6.44/7.03	1/7
					3	3	1	6.99/6.99	1/1
					3	3	2	7.42/7.42	14/7
					3	3	3	6.76/-	3/-

5 Conclusion

In the paper we presented the research concerning the identification of the important roles described on the basis of different characteristics of the activities in the blogosphere. Configurations of roles in groups and characteristics of such groups were analysed. The roles were considered both on the level of the whole network and for the given groups.

The performed analysis shows that it is possible to distinguish several classes of groups, considering the percentage of significant users belonging to them and their levels of cooperativeness. Groups with different sizes have various behaviour features caused by having a different percentage of influential users. Future works will concern the analysis of the lifespan of groups and identification of the core of groups - their parts that are strongly connected and most stable.

References

1. Asur, S., Parthasarathy, S., Ucar, D.: An event-based framework for characterizing the evolutionary behavior of interaction graphs. ACM Trans. Knowl. Discov. Data 3(4) (2009)
2. Fortunato, S.: Community detection in graphs. Phys. Rep., chap. 486 (2010)
3. Gleave, E., Welser, H., Lento, T., Smith, M.: A conceptual and operational definition of 'social role' in online community. In: 42nd Hawaii International Conference on System Sciences, HICSS 2009, pp. 1–11. IEEE (2009)

4. Gliwa, B., Koźlak, J., Zygmunt, A., Cetnarowicz, K.: Models of social groups in blogo-sphere based on information about comment addressees and sentiments. In: Aberer, K., Flache, A., Jager, W., Liu, L., Tang, J., Guéret, C. (eds.) SocInfo 2012. LNCS, vol. 7710, pp. 475–488. Springer, Heidelberg (2012)
5. Gliwa, B., Saganowski, S., Zygmunt, A., Bródka, P., Kazienko, P., Kozlak, J.: Identifi-cation of group changes in blogosphere. In: ASONAM 2012: IEEE/ACM Int. Conf. on Advances in Social Networks Analysis and Mining: Istanbul, Turkey (2012)
6. Hansen, D.L., Shneiderman, B., Smith, M.: Visualizing threaded conversation networks: Mining message boards and email lists for actionable insights. In: An, A., Lingras, P., Petty, S., Huang, R. (eds.) AMT 2010. LNCS, vol. 6335, pp. 47–62. Springer, Heidelberg (2010)
7. Junquero-Trabado, V., Dominguez-Sal, D.: Building a role search engine for social me-dia. In: Proc. of the 21st Int. Conf. Companion on World Wide Web, WWW 2012 Com-panio, pp. 1051–1060. ACM, NY (2012)
8. Mathioudakis, M., Koudas, N.: Efficient identification of starters and followers in social media. In: Proc. of the 12th Int. Conf. on Extending Database Technology: Advances in Database Technology, EDBT 2009, pp. 708–719. ACM, NY (2009)
9. Nolker, R., Zhou, L.: Social computing and weighting to identify member roles in online communities. In: Proceedings. The 2005 IEEE/WIC/ACM International Conference on Web Intelligence, pp. 87–93 (2005)
10. Palla, G., Derenyi, I., Farkas, I., Vicsek, T.: Uncovering the overlapping community structure of complex networks in nature and society. Nature 435, 814–818 (2005)
11. Spiliopoulou, M.: Evolution in social networks: A survey. In: Aggarwal, C.C. (ed.) Social Network Data Analytics. Springer (2011)
12. Wasserman, S., Faust, K.: Social Network Analysis: Methods and Applications. Cam-bridge University Press, Cambridge (1994)
13. Welser, H.T., Cosley, D., Kossinets, G., Lin, A., Dokshin, F., Gay, G., Smith, M.: Finding social roles in wikipedia. In: Proc. of the 2011 iConference, iConference 2011, pp. 122–129. ACM, New York (2011)
14. Zygmunt, A., Bródka, P., Kazienko, P., Kozlak, J.: Key person analysis in social commu-nities within the blogosphere. J. UCS 18(4), 577–597 (2012)